PROPERTIES OF A CENTER/SURROUND RETINEX
PART TWO: SURROUND DESIGN

Daniel J. Jobson Glenn A. Woodell
NASA Langley Research Center

ABSTRACT

The last version of Edwin Land's retinex model for human vision's lightness and color constancy has been implemented. Previous research has established the mathematical foundations of Land's retinex but has not examined specific design issues and their effects on the properties of the retinex operation. We have sought to define a practical implementation of the retinex without particular concern for its validity as a model for human lightness and color perception. Here we describe issues involved in designing the surround function. We find that there is a trade-off between rendition and dynamic range compression that is governed by the surround space constant. Various functional forms for the retinex surround are evaluated and a Gaussian form found to perform better than the inverse square suggested by Land. Preliminary testing led to the design of a Gaussian surround with a space constant of 80 pixels as a reasonable compromise between dynamic range compression and rendition.

1. Introduction

Of the many visual tasks accomplished so gracefully by human vision, one of the most fundamental and approachable for machine vision applications is lightness and color constancy. While a completely satisfactory definition is lacking, lightness and color constancy refer to human perception's resilience to wide ranging intensity and spectral illumination variations (scene-to-scene and to a large extent within scene). Various theories for this have been proposed and have a common mathematical foundation[1]. The last version of Edwin Land's retinex[2] has captured our attention because of the ease of implementation and manipulation of key variables, and because it does not have "unnatural" requirements for scene calibration. Likewise, the simplicity of the computation was appealing and initial experiments produced compelling results. This version of the retinex has been the subject of previous digital simulations that were limited because of lengthy computer

time involved and was implemented in analog VLSI to achieve real-time computation[3, 4]. Evidence that this retinex version is an optimal solution to the lightness problem has come from experiments posing Land's Mondrian target (randomly arranged two-dimensional gray patches) as a problem in linear optimization and a learning problem for back propagated artificial neural networks[5, 6].

The utility of a lightness-color constancy algorithm for machine vision is the simultaneous accomplishment of 1) dynamic range compression, 2) color independence from the spectral distribution of the scene illuminant, and 3) color and lightness rendition. Land's center/surround retinex demonstrably achieves the first two although Land emphasized primarily the color constancy properties. Well known difficulties arise though, for color and lightness rendition[1, 3, 6]. These consist of 1) lightness and color "halo" artifacts that are especially prominent where large uniform regions abut to form a high contrast edge with "graying" in the large uniform zones in an image, and 2) global violations of the gray world assumption (e.g., an all-red scene) which result in a global "graying out" of the image. Clearly the retinex (perhaps like human vision) functions best for highly diverse scenes and poorest for impoverished scenes. This is analogous to systems of simultaneous equations where a unique solution exists if and only if there are enough independent equations.

The general form of the center/surround retinex (Fig. 1) is similar to the difference-of-Gaussian (DOG) function widely used in natural vision science to model both the receptive fields of individual neurons and perceptual processes. The only extensions required are 1) to greatly enlarge and weaken the surround Gaussian (as determined by its space and amplitude constants), and 2) to include a logarithmic function to make subtractive inhibition into a shunting inhibition (i.e., arithmetic division). We have chosen a Gaussian surround form whereas Land opted for a $1/r^2$ function[2] and Moore et al.[3] used a different exponential form. These will be compared. Mathematically this takes the form,

$$R_i(x,y) = \log\ I_i(x,y) - \log\ [F(x,y) * I_i(x,y)] \tag{1}$$

where $I_i(x,y)$ is the image distribution in the ith color spectral band, "$*$" denotes the convolution operation, $F(x,y)$ is the surround function, and $R_i(x,y)$ is the associated retinex output. Equivalently

$$R_i(x,y) = \log\ I_i(x,y) - \log\ \left[\mathcal{F}^{-1}\{\hat{F}(\upsilon,\omega)\cdot\hat{I}_i(\upsilon,\omega)\}\right] \tag{2}$$

where $\hat{F}(\upsilon,\omega)$ and $\hat{I}_i(\upsilon,\omega)$ are the Fourier transforms of $F(x,y)$ and $I_i(x,y)$ and $\mathcal{F}^{-1}$ denotes the inverse Fourier transform.

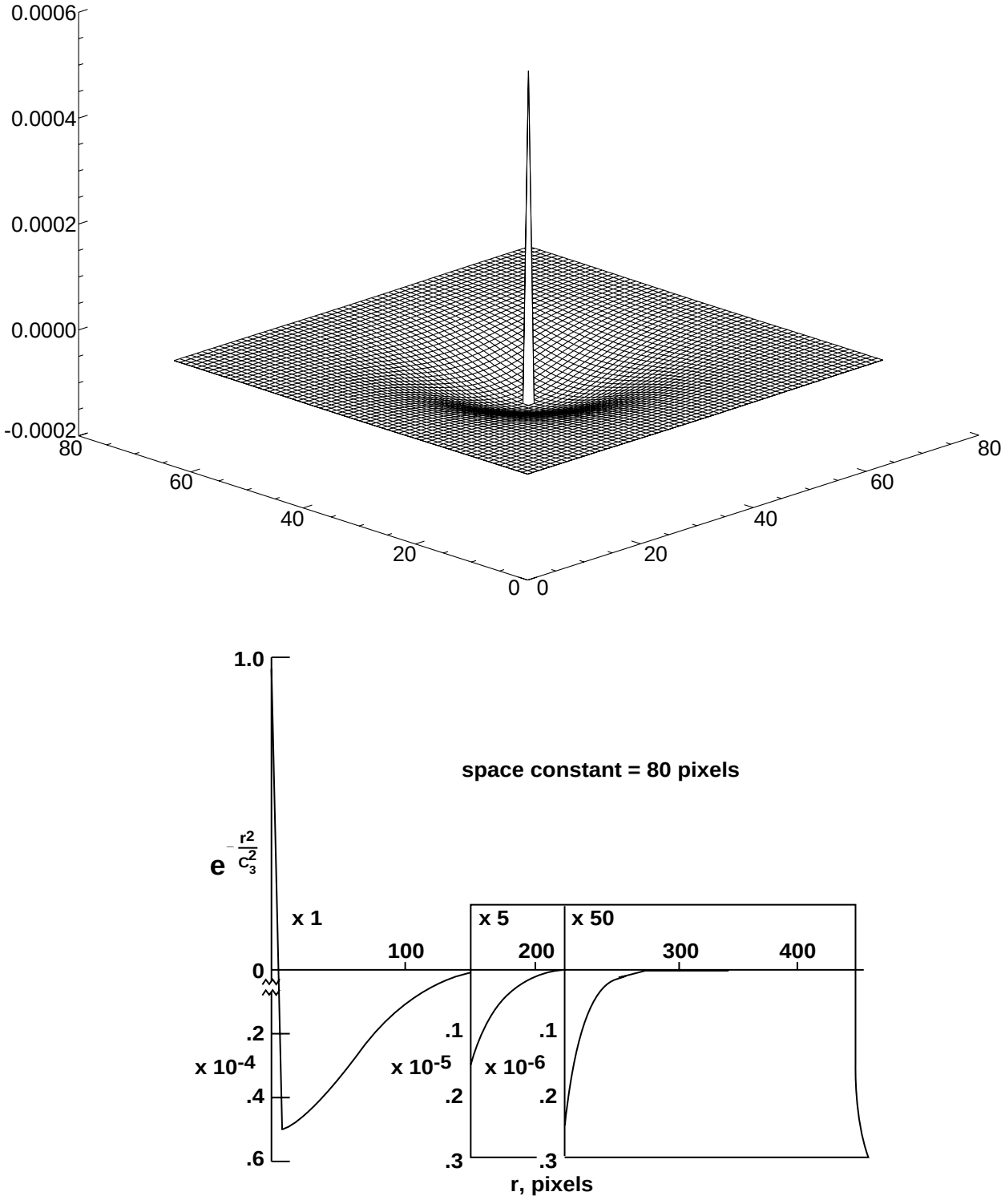

Figure 1: The spatial form of the center/surround retinex operator. a) 3-D representation (distorted to visualize surround). b) Cross-section to illustrate wide weak surround.

This operation is performed on each spectral band to produce Land's triplet values specifying color and lightness. It is readily apparent that color constancy (i.e., independence from single source illuminant spectral distribution) is reasonably complete since

$$I_i(x,y) = S_i(x,y)\ r_i(x,y) \tag{3}$$

where $S_i(x,y)$ is the spatial distribution of the source illumination and $r_i(x,y)$, the distribution of scene reflectances, so that

$$R_i(x,y) = \log\ \frac{S_i(x,y)\ r_i(x,y)}{\overline{S}_i(x,y)\ \overline{r}_i(x,y)} \tag{4}$$

where the bars denote the spatially weighted average value. As long as $S_i(x,y) \approx \overline{S}_i(x,y)$ then

$$R_i(x,y) \approx \log\ \frac{r_i(x,y)}{\overline{r}_i(x,y)}. \tag{5}$$

The approximate relation is an equality for many cases and, for those cases, where it is not strictly true, the reflectance ratio should dominate illumination variations.

This is demonstrated (Fig. 2) for the extreme cases of blue skylight illumination, direct sunlight only, and tungsten illumination. Actual daylight illumination should fall arbitrarily somewhere between the first two cases. Film and electronic cameras without computational intervention or film selection would produce the top row of images. Dynamic range compression is also readily demonstrated (Fig. 2(right)) with computer simulation. Here the original image data is multiplied by a hyperbolic tangent "shadow." Again cameras without computation produce the upper result (or with a change of f/stop or exposure would bring out the shadowed detail but at the expense of saturating the non-shadowed image zones). Strikingly, color balance is retained across the wide dynamic range encompassed and the highly nonlinear operation of the retinex.

The need for dynamic range compression and color constancy, especially if both are accomplished simultaneously by a simple real-time algorithm, is well-known to photographers. Discrepancies between the photographer's perception through the viewfinder and the captured film image can be quite bizarre (Fig. 3) and require constant vigilance to avoid impossible lighting situations, and to carefully select the appropriate film and processing for the illuminant's spectral distribution. The fundamental limit[3] is recognized to be the film or CRT's narrow dynamic range and static spectral response. Print/display dynamic range constraints of 50:1 are, however, compatible with the magnitude of scene reflectance variations. Except for extreme cases (snow or lampblack) reflectance variations are only 20:1[7] and often much less. Thus even the extremes of reflectance of $\approx$ 50:1 are easily spanned by print/display media. Clearly illumination variations are the culprit

Figure 2: Demonstration of retinex color constancy and dynamic range compression (prior to optimizing rendition) for a small space constant (space constant = 15 pixels).

which human visual perception has overcome by eye-brain computation. Electronic still and video cameras have an intrinsically high dynamic range ($> 2000{:}1$)[8] set by the detector array electronics and an even higher dynamic range within the detector array proper, since the limiting factor is usually the preamplifier noise added in transferring image signals off-chip or digitization noise added subsequently. Therefore, at least for electronic cameras, we can conclude that sufficient dynamic range is available to retain the full variations of both illumination and reflectance in arbitrary scenes. So it is certainly reasonable to consider either analog[3] implementations of compression/constancy or digital implementation if the initial A/D conversion is done at 10–14 bits, rather than the usual 8 bits.

Recent advances in high speed computing led us to reconsider both extensive digital simulations of the retinex and real time digital implementations for practical use in future electronic camera systems. The hours of computer time previously reported[3] are now reduced to minutes and real-time implementations using specialized digital hardware such as digital signal processing (DSP) chips are reasonable. In other words, the full image dynamic range is available from current electronic cameras, real time computation is realizable, and the ultimate bottleneck is only at the first print/display. Obviously, there are image coding aspects to both dynamic range compression and color constancy. Here, we will concentrate on the design of the algorithm to produce combined dynamic range compression/color constancy/color-lightness rendition.

We have seen that the center/surround retinex is both color constant and capable of a high degree of dynamic range compression. It remains then, to specify an implementation that produces satisfactory rendition and examine alternatives to determine if other design options are equally good or better. Because the retinex exchanges illumination variations for scene reflectance context dependency, scene content becomes a major issue especially when it deviates from regionally gray average values—the "gray world" assumption[1]. Therefore testing with diverse scenes, including random ones, is important to pinpoint possible limits to the generality of this retinex.

Initial image processing simulations revealed several unresolved implementation issues—(1) the placement of the log function, (2) the functional form of the surround, (3) the space constant for the surround, and (4) the treatment of the retinex triplets prior to display. These will be examined here while (1) and (4) are examined in a companion paper[9].

Figure 3: Examples of serious photographic defects due to spectral and/or spatial illumination variations. a) "Green" kitchen due to fluorescent illumination. b) Sodium vapor illumination. c) Tungsten indoors/daylight outdoors. d) Obscured foreground.

2. Surround Design Issues

A. The Surround Function

Land proposed an inverse square spatial surround

$$F(x,y) = 1/r^2 \tag{6}$$

where

$$r = \sqrt{x^2 + y^2}$$

which can be modified to be dependent on a space constant as

$$F'(x,y) = \frac{1}{1 + (r^2/c_1^2)}. \tag{7}$$

Moore[3] examined an exponential "absolute value"

$$F(x,y) = e^{-|r|/c_2} \tag{8}$$

because it is in approximation to the spatial response of analog VLSI resistive networks and Hurlbert[6] investigated the Gaussian

$$F(x,y) = e^{-r^2/c_3^2} \tag{9}$$

because of its widespread use in natural and machine vision modeling. A cross section of these two-dimensional functions (Fig. 4) shows that for any particular choice of space constant, the inverse square rolls-off very rapidly, but ultimately retains a higher response to quite distant image pixels than the exponential and Gaussian forms. At distant values, the exponential ultimately exceeds the Gaussian response, so that in general the inverse square is consistently more "global", the exponential is less so, and the Gaussian is more distinctively "regional."

In initial tests, no space constant for the inverse square surround could be found that achieved reasonable dynamic range compression, i.e., adequate enhancement of shadowed detail. The best performance is shown in Fig. 5. In contrast, both the exponential and Gaussian forms produced good dynamic range compression over a range of space constants. Because the Gaussian offered the most experimental flexibility (good performance over wider range of space constants), it was selected for this implementation. It is likely that the exponential is equally useful and this is clearly of importance for analog VLSI resistive network hardware implementations of retinex computations.

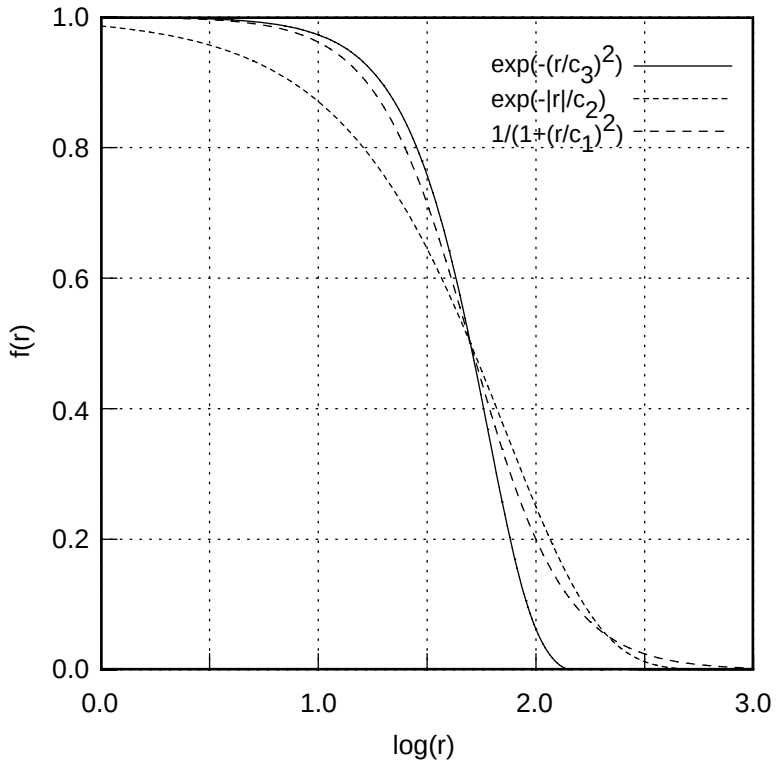

Figure 4: Comparison of three surround functions—inverse square, exponential, and Gaussian, normalized to equal full-width half-max (FWHM) response. The $\log(r)$ scale is necessary for comparison purposes but does diminish the differences between the functions. A linear r scale (if it were graphically feasible) would show very dramatic differences. The space constants are $c_1 = 50$ pixels, $c_2 = 72$ pixels, and $c_3 = 60$ pixels.

Figure 5: Comparison of visual performance of three surround functions arranged from left to right in order of increasing dynamic range compression.

B. Surround Space Constant

While Land proposed the center/surround retinex with a 2–4 pixel diameter for the center (perhaps in keeping with the widely known coarser spatial resolution of purely chromatic vision), a center of only one pixel is clearly demanded for general purpose image processing. Only after segmentation into lightness and chromatic images can the purely chromatic images be made coarser. In contrast, the surround space constant cannot be so clearly defined. Land proposed an inverse square surround with a full width-half maximum (FWHM) of 40° of visual angle. This corresponds to FWHM of about 270 visual pixels (assuming a visual pixel is $\approx 0.015°$). We examined the performance of the Gaussian surround over a wide range of space constants. Since previous research[6] found variations in the space constant with the spatial variation in shadow profiles, a particular concern is the question of an optimum space constant that gives good performance for diverse scenes and lighting conditions.

The image sequence (Fig. 6) established a trade-off which has not been previously studied. In varying the space constant from small to large values, dynamic range compression is sacrificed for improved rendition. The middle of this range ($50 \leq c_3 \leq 100$ pixels) represents a reasonable compromise where shadows are fairly compensated and rendition achieves acceptable levels of image quality. This is qualitatively compatible with human visual perception in that the treatment of shadows is influenced by their spatial extent. Larger shadows tend to be more compensated (less dark) while smaller shadows appear less compensated (blacker and with less visible internal detail).

While we are not concerned with defining a form of the retinex which accurately models human vision, we must ultimately compare performance to that of human perception in order to meet basic image quality requirements. Our intent then is to find a form of the retinex which is functionally equivalent to human visual perception. Since the performance of human vision for complex natural images has not been comprehensively defined, we are left with purely subjective assessments of image quality.

3. Conclusions

The specific surround implementation we have defined from preliminary testing is an operation with the following characteristics:

1. The form of the surround is Gaussian.

Figure 6: The tradeoff between dynamic range compression and color rendition for the Gaussian surround. Small space constants produce excellent dynamic range compression while large constants produce the best rendition.

2. The spatial extent of the surround is that for a Gaussian space constant of about 80 pixels (which corresponds to a FWHM spread of 210 pixels).

In a companion paper,[9] additional components of the design are defined:

3. The logarithm is applied after surround formation by two-dimensional spatial convolution.
4. A "canonical" gain-offset is applied to the retinex output which, in signal terms, clips some of the highest and lowest signal excursions.

REFERENCES

1. A. C. Hurlbert, "Formal connections between lightness algorithms," *Journal of the Optical Society of America A*, vol. 3, pp. 1684–1693, 1986.

2. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3078–3080, 1986.

3. A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," *IEEE Transactions on Neural Networks*, vol. 2, pp. 237–247, March 1991.

4. A. Moore, G. Fox, J. Allman, and R. M. Goodman, "A VLSI neural network for color constancy," in *Advances in Neural Information Processing 3* (D. S. Touretzky and R. Lippman, eds.), pp. 370–376, San Mateo, CA: Morgan Kaufmann, 1991.

5. A. C. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science*, vol. 239, pp. 482–485, 1988.

6. A. C. Hurlbert, *The Computation of Color*. PhD thesis, Massachusetts Institute of Technology, September 1989.

7. D. E. Bowker, R. E. Davis, D. L. Myrick, K. Stacy, and W. L. Jones, "Spectral reflectances of natural targets for use in remote sensing studies," *NASA Reference Publication* 1139, June 1985.

8. R. H. Dyck, "Design, fabrication, and performance of ccd imagers," in *VLSI Electronics Microstructures Science* (N. G. Einspruch, ed.), vol. 3, pp. 65–107, Orlando, FL: Academic Press, 1982.

9. Z. Rahman, "Properties of a center/surround Retinex Part One: Signal processing design," *NASA Contractor Report* 198194, 1995.